1 OPTIMAL ESTIMATION

1.1 Variance of a Random Variable

1.2 Estimation Given No Observations

1.3 Estimation Given Dependent Observations

1.3.1 Mean-Square-Error Criterion

1.3.2 Orthogonality Principle

1.3.3 Gaussian Random Variables

1.4 Estimation in the Complex and Vector Cases

1.4.1 Complex-Valued Random Variables

1.4.2 Vector-Valued Random Variables

1.4.3 Optimal Estimator in the Vector Case

1.4.4 Equivalent Optimization Criterion

1.4.5 Spherically Invariant Gaussian Variables

1.5 Summary of Main Results

1.6 Bibliographic Notes

1.7 Problems

1.8 Computer Project

1.A Hermitian and Positive-Definite Matrices

1.B Gaussian Random Vectors

1.1 Variance of a Random Variable

WMU Undergraduate class ECE 3800

Random variable: A real function whose domain is that of the outcomes of an experiment (sample space, S) and whose actual value is unknown in advance of the experiment.

From: http://en.wikipedia.org/wiki/Random_variable

A random variable can be thought of as the numeric result of operating a non-deterministic mechanism or performing a non-deterministic experiment to generate a random result.

Unlike the common practice with other mathematical variables, a random variable cannot be assigned a value; a random variable does not describe the actual outcome of a particular experiment, but rather describes the possible, as-yet-undetermined outcomes in terms of real numbers.

Probability Distribution Function (PDF) or Cumulative Distribution Function (CDF)

Probability Distribution Function:The probability of the event that the observed random variable *X* is less than or equal to the allowed value x.

$$F_X(x) = \Pr(X \leq x)$$

The defined function can be discrete or continuous along the x-axis. Constraints on the probability distribution function are:

1. $0 \leq F_X(x) \leq 1, \; for -\infty < x < \infty$
2. $F_X(-\infty) = 0$ and $F_X(\infty) = 1$
3. F_X is non-decreasing as x increases
4. $\Pr(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$

Note: This is also known as the Cumulative Distribution function or CDF in many references! (this avoids a major problems with language that is coming soon …)

From: http://en.wikipedia.org/wiki/Cumulative_distribution_function

In probability theory, the cumulative distribution function (abbreviated cdf) completely describes the probability distribution of a real-valued random variable, X. For every real number x, the cdf is given by

$$F_X(x) = \Pr(X \leq x)$$

where the right-hand side represents the probability that the variable X takes on a value less than or equal to x. The probability that X lies in the interval (a, b) is therefore F(b) − F(a) if a ≤ b. It is conventional to use a capital *F* for a cumulative distribution function, in contrast to the lower-case *f* used for *probability density functions* and *probability mass functions*.

Probability Density Function (pdf)

The derivative of the probability distribution function

$$f_X(x) = \lim_{\varepsilon \to 0} \frac{F_X(x+\varepsilon) - F_X(x)}{\varepsilon} = \frac{dF_X(x)}{dx}$$

An interpretation is

$$f_X(x) \cdot dx = \Pr(x < X \le x + dx) = F_X(x+dx) - F_X(x)$$

Properties of the pdf include

1. $f_X(x) \ge 0, \; for - \infty < x < \infty$

2. $\int_{-\infty}^{\infty} f_X(x) \cdot dx = 1$

3. $F_X(x) = \int_{-\infty}^{x} f_X(u) \cdot du$

4. $\Pr(x_1 < X \le x_2) = \int_{x_1}^{x_2} f_X(x) \cdot dx$

From: http://en.wikipedia.org/wiki/Probability_density_function

In mathematics, a probability density function (pdf) serves to represent a probability distribution in terms of integrals. A probability density function is everywhere non-negative and its integral from −∞ to +∞ is equal to 1. If a probability distribution has density *f(x)*, then intuitively the infinitesimal interval *[x, x + dx]* has probability *f(x) dx*.

Mean Values and Moments

Mean Value: the expected mean value of measurements of a process involving a random variable.

This is commonly called the expectation operator or expected value of …
and is mathematically described as:

$$\overline{X} = E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx$$

For laboratory experiments, the expected value of a voltage measurement can be thought of as the DC voltage.

In general, the expected value of a function is:

$$E[g(X)] = \int_{-\infty}^{\infty} g(X) \cdot f_X(x) \cdot dx$$

Moments

The moments of a random variable are defined as the expected value of the powers of the measured output or …

$$\overline{X^n} = E[X^n] = \int_{-\infty}^{\infty} x^n \cdot f_X(x) \cdot dx$$

Therefore, the mean or average is sometimes called the first moment.

Expected Mean Squared Value or Second Moment

The mean square value or second moment is

$$\overline{X^2} = E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x) \cdot dx$$

Central Moments

The central moments are the moments of the difference between a random variable and its mean.

$$\overline{\left(X - \overline{X}\right)^n} = E\left[\left(X - \overline{X}\right)^n\right] = \int_{-\infty}^{\infty} \left(x - \overline{X}\right)^n \cdot f_X(x) \cdot dx$$

Notice that the first central moment is ***0*** …

The second central moment is referred to as the variance of the random variable …

$$\sigma^2 = \overline{\left(X - \overline{X}\right)^2} = E\left[\left(X - \overline{X}\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \overline{X}\right)^2 \cdot f_X(x) \cdot dx$$

Note that:

$$\sigma^2 = E\left[\left(X - \overline{X}\right)^2\right] = E\left[\left(X - \overline{X}\right) \cdot \left(X - \overline{X}\right)\right]$$

$$\sigma^2 = E\left[X^2 - 2 \cdot X \cdot \overline{X} + \overline{X}^2\right]$$

$$\sigma^2 = E\left[X^2\right] - 2 \cdot \overline{X} \cdot E[X] + \overline{X}^2$$

$$\sigma^2 = E\left[X^2\right] - 2 \cdot \overline{X} \cdot \overline{X} + \overline{X}^2$$

$$\sigma^2 = E\left[X^2\right] - \overline{X}^2 = E\left[X^2\right] - E[X]^2$$

$$\sigma^2 = \overline{X^2} - \overline{X}^2$$

A small variance indicates that a random variable is more likely to assume values that are close to the mean.

A large variance indicates that a random variable is more likely to assume a wider interval of values that are around the mean.

Therefore, it describes a measure of uncertainty.

Gaussian or Normal probability density function

The Gaussian or Normal probability density function is defined as:

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(\frac{-\left(x - \overline{X}\right)^2}{2 \cdot \sigma^2} \right), \; for \; -\infty < x < \infty$$

where $\overline{X}$ is the mean and σ is the variance

Reasons for importance:

1. It provides a good mathematical model for a great many different physically observed random phenomena. Furthermore, the fact that it should be a good model can be justified theoretically in many ways.

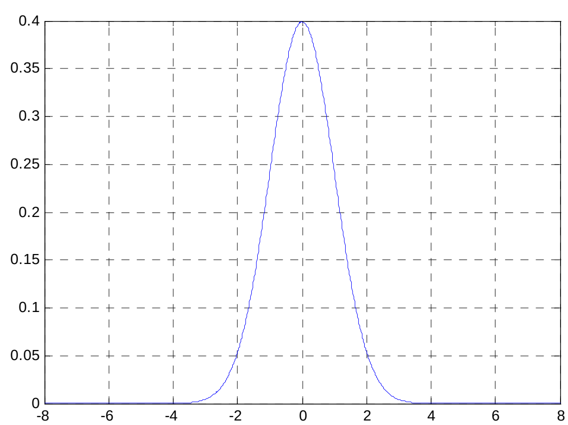
2. It is one of the few density functions that can be extended to handle an arbitrarily large number of random variables conveniently.

3. Linear combinations of Gaussian random variables lead to new random variables that are also Gaussian. This is not true for most other density functions.

4. The random process from which Gaussian random variables are derived can be completely specified, in a statistical sense, from knowledge of all first and second moments only. This is not true for other processes.

5. In system analysis, the Gaussian process is often the only one for which a complete statistical analysis can be carried through in either the linear or nonlinear situation.

6. The function is infinitely differentiable (all the derivatives exist).



Important notes on the curve:

1. There is only one maximum and it occurs at the mean value.

2. The density function is symmetric about the mean value.

3. The width of the density function is directly proportional to the standard deviation, σ. The width of 2σ occurs at the points where the height is 0.607 of the maximum value. These are also the points of the maximum slope.

4. The maximum value of the density function is inversely proportional to the standard deviation, σ. Since the density function has an area of unity, it can be used as a representation of the impulse or delta function by letting σ approach zero. That is

$$\delta\left(x-\overline{X}\right)=\lim_{\sigma\to 0}\left[\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(\frac{-\left(x-\overline{X}\right)^2}{2\cdot\sigma^2}\right)\right]$$

The Gaussian Probability Distribution Function is

$$F_X(x)=\int_{v=-\infty}^{x}\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\exp\left(\frac{-\left(v-\overline{X}\right)^2}{2\cdot\sigma^2}\right)\cdot dv$$

The PDF can not be represented in a closed form solution!

The PDF is tabulated in ***multiple textbooks*** for a zero mean, unit variance pdf.
For these values, it is often described as “normalized” and is defined as

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_{u=-\infty}^{x}\exp\left(-\frac{u^2}{2}\right)\cdot du$$

The distribution function is then defined as

$$F_X(x)=\Phi\left(\frac{x-\overline{X}}{\sigma}\right)$$

When using a textbook table, the negative values in x are derived as

$$\Phi(-x)=1-\Phi(x)$$

Another defined function that is related to the Gaussian (and used) is the ***Q-function***,

$$Q(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_{u=x}^{\infty}\exp\left(-\frac{u^2}{2}\right)\cdot du$$

The ***Q-function*** is the complement of the normal function, Φ:

$$Q(x)=1-\Phi(x)$$

Therefore note that:

$$Q(-x)=1-Q(x)$$

$$F_X(x)=1-Q\left(\frac{x-\overline{X}}{\sigma}\right)$$

An approximation for the Q function

$$Q(a)\cong\left(1-\frac{1}{2\cdot a^2}\right)\cdot\frac{1}{a\cdot\sqrt{2\pi}}\cdot\exp\left(-\frac{a^2}{2}\right),\qquad for\ x=a>3$$

Another way to find values for the Gaussian

The error function

$$erf(x) = \frac{2}{\sqrt{\pi}} \cdot \int_{u=0}^{x} \exp\left(-u^2\right) \cdot du$$

$$Q(x) = \frac{1}{2} \cdot \left[1 - erf\left(\frac{x}{\sqrt{2}}\right)\right]$$

$$F_X(x) = 1 - \frac{1}{2} \cdot \left[1 - erf\left(\frac{x - \overline{X}}{\sqrt{2} \cdot \sigma}\right)\right] = \frac{1}{2} + \frac{1}{2} \cdot erf\left(\frac{x - \overline{X}}{\sqrt{2} \cdot \sigma}\right)$$

The error function (`Y = ERF(X)`) is built-in to MATLAB. ***Appendix G*** provides an overview of the functions and how they can be used within MATLAB.

Note: There is a typo in the definition for the ***erf*** on p. 441. The integration should be as shown above, not x to infinity.

From MATLAB:

```
ERF Error function.
    Y = ERF(X) is the error function for each element of X.  X must be
    real. The error function is defined as:

      erf(x) = 2/sqrt(pi) * integral from 0 to x of exp(-t^2) dt.

    See also erfc, erfcx, erfinv.

    Reference page in Help browser
       doc erf
```

Rayleigh Distribution

For a two dimensional problem (positions in x and y with two independent Gaussian random variable noise or offset terms), the distance from a desired point is described as a radial or vector magnitude, the radial error or offset is described by the Rayleigh Distribution.

For $R = \sqrt{X^2 + Y^2}$

The probability density function (pdf) is

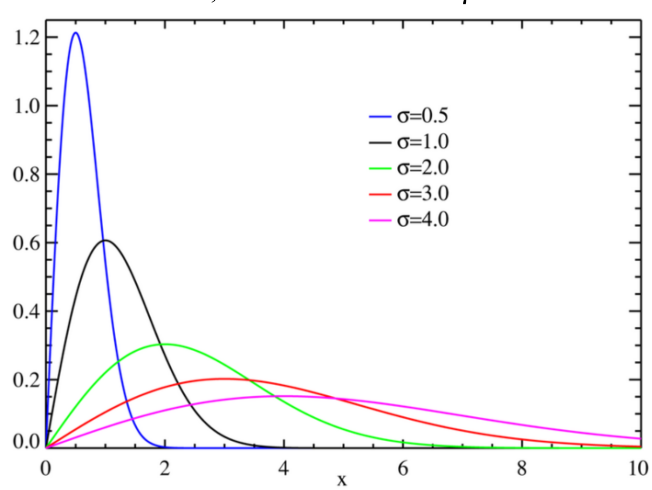
$$\begin{aligned} f_R(r) &= \frac{r}{\sigma^2} \cdot \exp\left(\frac{-r^2}{2 \cdot \sigma^2}\right), && for\ 0 \le r \\ &= 0, && for\ r < 0 \end{aligned}$$

The probability distribution function (PDF) can be derived as:

$$\begin{aligned} F_R(r) &= \int_{v=0}^{r} \frac{v}{\sigma^2} \cdot \exp\left(\frac{-v^2}{2 \cdot \sigma^2}\right) \cdot dv, && for\ 0 \le r \\ &= 0, && for\ r < 0 \end{aligned}$$

$$\begin{aligned} F_R(r) &= -\exp\left(\frac{-r^2}{2 \cdot \sigma^2}\right)\Bigg|_0^r, && for\ 0 \le r \\ &= 0, && for\ r < 0 \end{aligned}$$

$$\begin{aligned} F_R(r) &= 1 - \exp\left(\frac{-r^2}{2 \cdot \sigma^2}\right), && for\ 0 \le r \\ &= 0, && for\ r < 0 \end{aligned}$$



http://en.wikipedia.org/wiki/Rayleigh_distribution

The first moment

$$\overline{R} = E[R] = \int_{r=0}^{\infty} r \cdot \frac{r}{\sigma^2} \cdot \exp\left(\frac{-r^2}{2 \cdot \sigma^2}\right) \cdot dr$$

Closed form solution (p. 421) $\displaystyle\int_{r=0}^{\infty} x^2 \cdot \exp\left(-a^2 \cdot x^2\right) \cdot dx = \frac{\sqrt{\pi}}{4 \cdot a^3}$

$$\overline{R} = E[R] = \frac{1}{\sigma^2} \cdot \frac{\sqrt{\pi}}{4} \cdot \left(\sqrt{2} \cdot \sigma\right)^3 = \frac{\sqrt{\pi}}{\sqrt{2}} \cdot \sigma = \sqrt{\frac{\pi}{2}} \cdot \sigma$$

The second moment

$$\overline{R^2} = E\left[R^2\right] = \int_{r=0}^{\infty} r^2 \cdot \frac{r}{\sigma^2} \cdot \exp\left(\frac{-r^2}{2 \cdot \sigma^2}\right) \cdot dr$$

$$\overline{R^2} = E\left[R^2\right] = 2 \cdot \sigma^2$$

The second central moment, variance or standard deviation is

$$E\left[\left(r - \overline{R}\right)^2\right] = \sigma_R^2 = \overline{R^2} - E[R]^2 = 2 \cdot \sigma^2 - \left(\sqrt{\frac{\pi}{2}} \cdot \sigma\right)^2$$

$$E\left[\left(r - \overline{R}\right)^2\right] = \sigma_R^2 = \overline{R^2} - E[R]^2 = \left(2 - \frac{\pi}{2}\right) \cdot \sigma^2$$

Chebyshev Inequality

$$\Pr\left(\left|x - E[x]\right| \geq \delta\right) \leq \frac{\sigma_x^2}{\delta^2}$$

or

$$\Pr\left(\left|x - E[x]\right| \geq \delta \cdot \sigma_x\right) \leq \frac{1}{\delta^2}$$

If you know the variance, you can make an estimate of the probability that a value will be outside the two-sided interval defined by $(E[x] - \delta, E[x] + \delta)$

For a zero-mean Gaussian random variable, what would the inequality predict for values outside twice the variance?

$$\Pr\left(|x| \geq \delta\right) \leq \frac{\sigma_x^2}{\delta^2}$$

$$\delta = 2 \cdot \sigma_x$$

$$\Pr\left(|x| \geq 2 \cdot \sigma_x\right) \leq \frac{\sigma_x^2}{4 \cdot \sigma_x^2} = 25\%$$

This is not a very tight bound, but a simple rapid approximation.

Using an actual normal probability distribution table, the actual value is

$$\Pr\left(|x| \geq 2 \cdot \sigma_x\right) \leq 2 \cdot \left(1 - \Phi(2.0)\right) = 2 \cdot (1 - 0.9772) = 4.56\%$$

What if the variance is zero? Chebyshev Inequality says …

$$\Pr\left(\left|x - E[x]\right| \geq \delta\right) \leq \frac{0}{\delta^2} = 0$$

Thus the probability that the random variable and its mean are different, by any amount is zero!

1.2 Estimation Given No Observations

Mean square error criteria: Solving for an estimate x by minimizing the mean-square-error (mse). This results in the minimum mse or mmse.

The mean of x

$$\bar{x} = E[x]$$

The error signal given an estimate of x is defined as

$$\tilde{x} \equiv x - \hat{x}$$

where $\hat{x}$ is the optimal estimate of x in a mean square sense. That is to say $\hat{x}$ is the value that minimizes

$$\min_{\hat{x}}\left\{E\left[\tilde{x}^2\right]\right\} = \min_{\hat{x}}\left\{E\left[(x-\hat{x})^2\right]\right\}$$

Proof – for lack of any observations

$$E\left[\tilde{x}^2\right] = E\left[(x-\hat{x})^2\right] = E\left[(x-\bar{x}+\bar{x}-\hat{x})^2\right]$$

$$E\left[\tilde{x}^2\right] = E\left[(x-\bar{x})^2 + 2\cdot(x-\bar{x})\cdot(\bar{x}-\hat{x}) + (\bar{x}-\hat{x})^2\right]$$

$$E\left[\tilde{x}^2\right] = \sigma_x^2 + 2\cdot(\bar{x}-\hat{x})\cdot E[(x-\bar{x})] + (\bar{x}-\hat{x})^2$$

$$E\left[\tilde{x}^2\right] = \sigma_x^2 + (\bar{x}-\hat{x})^2$$

Therefore to minimize without observations, the mmse estimate must be

$$\hat{x} = \bar{x}$$

and the mmse variance is

$$E\left[\tilde{x}^2\right] = \sigma_x^2 = E\left[(x-\bar{x})^2\right]$$

Define a cost function, J, based on the error mean and variance

$$J[\tilde{x}] = A\cdot E\left[\tilde{x}^2\right] + B\cdot E[\tilde{x}]$$

When the mmse is selected,

$$J[\tilde{x}] = A\cdot\sigma_x^2 + B\cdot E[x-\hat{x}] = A\cdot\sigma_x^2$$

1.3 Estimation Given Dependent Observations

(A review of joint and conditional probability)

Joint Probability Distribution Function (PDF)

Probability Distribution Function:The probability of the event that the observed random variable X is less than or equal to the allowed value x.

$$F_{XY}(x,y) = \Pr(X \le x, Y \le y)$$

The defined function can be discrete or continuous along the x-axis. Constraints on the probability distribution function are:

1. $0 \le F_{XY}(x,y) \le 1, \qquad for -\infty < x < \infty \ and \ -\infty < y < \infty$
2. $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = F_{XY}(-\infty, -\infty) = 0$
3. $F_{XY}(\infty, \infty) = 1$
4. $F_{XY}(x,y)$ is non-decreasing as either x or y increases
5. $F_{XY}(x,\infty) = F_X(x)$ and $F_{XY}(\infty, y) = F_Y(y)$

Joint Probability Density Function (pdf)

The derivative of the probability distribution function

$$f_{XY}(x,y) = \frac{\partial^2 F_X(x)}{\partial x \partial y}$$

Properties of the pdf include

1. $f_{XY}(x,y) \ge 0, \quad for -\infty < x < \infty \ and \ -\infty < y < \infty$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y) \cdot dx \cdot dy = 1$

Note: the volume of the 2-D density function is one.

3. $F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(u,v) \cdot du \cdot dv$
4. $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y) \cdot dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y) \cdot dx$
5. $\Pr(x_1 < X \le x_2, y_1 < Y \le y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x,y) \cdot dx \cdot dy$

Conditional Probability (Again, with multiple r.v.)

Using the Probability Distribution Function (PDF), define

$$F_X(x \mid Y \le y) = \frac{\Pr(X \le x \mid M)}{\Pr(M)} = \frac{F_{XY}(x,y)}{F_Y(y)}$$

Another way.

$$F_X(x \mid y_1 \le Y \le y_2) = \frac{F_{XY}(x,y_2) - F_{XY}(x,y_1)}{F_Y(y_2) - F_Y(y_1)}$$

And finally,

$$F_X(x \mid Y = y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

also

$$F_Y(y \mid X = x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

these are special case as for the pdf in X and Y…

$$f_X(X = x) = 0 \quad \text{and} \quad f_Y(Y = y) = 0$$

An *engineering* derivation follows:

$$F_X(x \mid Y = y) = \lim_{\Delta y \to 0} \frac{F_{XY}(x, y+\Delta y) - F_{XY}(x,y)}{F_Y(y+\Delta y) - F_Y(y)} = \lim_{\Delta y \to 0} \frac{\left[F_{XY}(x, y+\Delta y) - F_{XY}(x,y)\right] / \Delta y}{\left[F_Y(y+\Delta y) - F_Y(y)\right] / \Delta y}$$

$$F_X(x \mid Y = y) = \frac{\partial F_{XY}(x,y) / \partial y}{\partial F_Y(y) / \partial y} = \frac{\int_{-\infty}^{x} f_{XY}(u,y) \cdot du}{f_Y(y)}$$

The corresponding pdf is (taking the derivative with respect to x

$$f_{XY}(x \mid Y = y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \text{and alternately} \quad f_{XY}(y \mid X = x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

Bayes Rule can then be stated as

$$f_{XY}(x,y) = f_{XY}(x \mid Y = y) \cdot f_Y(y) = f_{XY}(y \mid X = x) \cdot f_X(x)$$

The ***total probability*** is stated using:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y) \cdot dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y) \cdot dx$$

resulting in

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y) \cdot dx = \int_{-\infty}^{\infty} f_{XY}(y \mid X = x) \cdot f_X(x) \cdot dx$$

or

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y) \cdot dy = \int_{-\infty}^{\infty} f_{XY}(x \mid Y = y) \cdot f_Y(y) \cdot dy$$

To derive the multiple variable ***Bayes Theorem***, use

$$f_{XY}(x,y) = f_{XY}(x \mid Y = y) \cdot f_Y(y) = f_{XY}(y \mid X = x) \cdot f_X(x)$$

resulting in

$$f_{XY}(x \mid Y = y) = \frac{f_{XY}(x,y) \cdot}{f_Y(y)}$$

or

$$f_{XY}(y \mid X = x) = \frac{f_{XY}(x,y) \cdot}{f_X(x)}$$

Note: the joint probability density function completely specifies:

- both marginal density functions and
- both conditional density functions.

Statistical Independence

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

Definition of correlation for X and Y independent

$$E[X \cdot Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{XY}(x,y) \cdot dx \cdot dy = \left[\int_{-\infty}^{\infty} x \cdot f_X(x) \cdot dx\right] \cdot \left[\int_{-\infty}^{\infty} y \cdot f_Y(y) x \cdot dy\right]$$

$$E[X \cdot Y] = E[X] \cdot E[Y] = \overline{X} \cdot \overline{Y}$$

As another consequence

$$f_{XY}(x \mid Y = y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_Y(y) \cdot f_X(x)}{f_Y(y)} = f_X(x)$$

and similarly

$$f_{XY}(y \mid X = x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_X(x) \cdot f_Y(y)}{f_X(x)} = f_Y(y)$$

Correlation between Random Variables

$$E[X \cdot Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot y \cdot f_{XY}(x,y) \cdot dx \cdot dy$$

Developing values where the random variable means have been extracted, the ***covariance***

$$\sigma_{XY} = E[(X - E[X]) \cdot (Y - E[Y])] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X) \cdot (y - \mu_Y) \cdot f_{XY}(x,y) \cdot dx \cdot dy$$

This gives rise to another factor, when the random variable variance is also used to normalize the factors as:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right) \cdot \left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{x - \mu_X}{\sigma_X}\right) \cdot \left(\frac{y - \mu_Y}{\sigma_Y}\right) \cdot f_{XY}(x,y) \cdot dx \cdot dy$$

This equation defines the ***correlation coefficient*** or ***normalized covariance***;
the modified random variables are called the ***standardized variables*** and have zero mean and a unit variance.

An alternate expression for the correlation coefficient is

$$\rho = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{x \cdot y - \mu_X \cdot y - \mu_Y \cdot x + \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y}\right) \cdot f_{XY}(x,y) \cdot dx \cdot dy$$

$$\rho = \frac{E[x \cdot y] - \mu_X \cdot \mu_Y}{\sigma_X \cdot \sigma_Y}$$

For random variables that are inherently zero mean with a unit variance,

$$\rho = E[x \cdot y]$$

getting back to the text …..

1.3 <u>Estimation</u> Given Dependent Observations

When we have observations of a random variable in a noise environment, similar to estimating the voltage of a signal in noise.

$$y = x + v$$

where x and v are random variables. Determine the least mean square estimator of x given y.

$$\hat{x} = h(y)$$

For each measured value of y, a different estimate may exist as

$$\hat{x} = h(y)_{y=y} = h(y)$$

(note: notation appears wrong, but check text for random variable or measured value notation)

The error signal given an estimate of x is defined as

$$\tilde{x} \equiv x - \hat{x}$$

where $\hat{x}$ is the optimal estimate of x in a mean square sense. That is to say $\hat{x}$ is the value that minimizes

$$\min_{h(\cdot)}\left\{E\left[\tilde{x}^2\right]\right\} = \min_{h(\cdot)}\left\{E\left[(x-\hat{x})^2\right]\right\}$$

Determining the least-mean-square-estimator (lmse) of x given y

$$\hat{x} = E[x \mid y] = \int_{S_x} x \cdot f_{XY}(x \mid y) \cdot dx$$

where the integral denotes the domain of x.

This estimator is unbiased as

$$\hat{x} = E[x] = \bar{x} = \int_{S_x} x \cdot f_X(x) \cdot dx$$

The proof is in the text and reviewed in homework problem 1.4.

The minimum cost or variance becomes

$$E\left[\tilde{x}^2\right] = E\left[x^2 - \hat{x}^2\right] = \sigma_x^2 - \sigma_{\hat{x}}^2$$

Example 1.3.1 Noisy Measurement of a binary signal

Binary values for transmission take on the values of +/- 1 with a desired probability of ½ (not always true, but usually desired).

If we make observation/estimate in a zero-mean, unit variance Gaussian environment:

$$y = x + v$$

the probability density of y (based on x and v independent) is

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x) \cdot f_V(y-x) \cdot dx$$

$$f_X(x) = \frac{\delta(x-1)}{2} + \frac{\delta(x+1)}{2} \qquad f_V(v) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(\frac{-v^2}{2}\right)$$

$$f_Y(y) = \int_{-\infty}^{\infty} \left(\frac{\delta(x-1)}{2} + \frac{\delta(x+1)}{2}\right) \cdot f_V(y-x) \cdot dx$$

$$f_Y(y) = \frac{f_V(y-1)}{2} + \frac{f_V(y+1)}{2}$$

The joint density of x and y is

$$f_{XY}(x,y) = f_X(x) \cdot f_{XY}(y \mid x)$$

Note that y|x defines the value of v, a random variable independent of x

$$f_{XY}(x,y) = f_X(x) \cdot f_V(y-x)$$

The estimate of x given y from before is defined as

$$\hat{x} = E[x \mid y] = \int_{S_x} x \cdot f_{XY}(x \mid y) \cdot dx$$

$$f_{XY}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)}$$

$$f_{XY}(x \mid y) = \frac{f_X(x) \cdot f_V(y-x)}{\frac{f_V(y-1)}{2} + \frac{f_V(y+1)}{2}} = \frac{\left(\frac{\delta(x-1)}{2} + \frac{\delta(x+1)}{2}\right) f_V(y-x)}{\frac{f_V(y-1)}{2} + \frac{f_V(y+1)}{2}}$$

$$f_{XY}(x \mid y) = \frac{(\delta(x-1)+\delta(x+1))f_V(y-x)}{f_V(y-1)+f_V(y+1)} = \frac{\delta(x-1)\cdot f_V(y-x)}{f_V(y-1)+f_V(y+1)} + \frac{\delta(x+1)\cdot f_V(y-x)}{f_V(y-1)+f_V(y+1)}$$

Then

$$\hat{x} = E[x \mid y] = \int_{S_x} x \cdot \left[\frac{\delta(x-1)\cdot f_V(y-x)}{f_V(y-1)+f_V(y+1)} + \frac{\delta(x+1)\cdot f_V(y-x)}{f_V(y-1)+f_V(y+1)} \right] \cdot dx$$

$$\hat{x} = E[x \mid y] = \frac{f_V(y-1)}{f_V(y-1)+f_V(y+1)} - \frac{f_V(y+1)}{f_V(y-1)+f_V(y+1)}$$

$$\hat{x} = E[x \mid y] = \frac{1}{1+\dfrac{f_V(y+1)}{f_V(y-1)}} - \frac{1}{\dfrac{f_V(y-1)}{f_V(y+1)}+1}$$

$$\frac{f_V(y+1)}{f_V(y-1)} = \frac{\dfrac{1}{\sqrt{2\pi}}\cdot \exp\left(\dfrac{-(y+1)^2}{2}\right)}{\dfrac{1}{\sqrt{2\pi}}\cdot \exp\left(\dfrac{-(y-1)^2}{2}\right)} = \frac{\exp\left(\dfrac{-2\cdot y}{2}\right)}{\exp\left(\dfrac{+2\cdot y}{2}\right)} = \frac{\exp(-y)}{\exp(+y)}$$

$$\hat{x} = E[x \mid y] = \frac{\exp(+y)}{\exp(+y)+\exp(-y)} - \frac{\exp(-y)}{\exp(+y)+\exp(-y)} = \frac{\exp(+y)-\exp(-y)}{\exp(+y)+\exp(-y)} = \tanh(y)$$

$$\hat{x} = \tanh(y)$$

Note that the estimator does not provide a "logical value" for the bit transmitted, but a value based on a functions where the maxima and minim are +1 and -1.

Since it is known a-priori that x takes on a unit magnitude, it is logical to make the bit value selected for x using a modified estimator as

$$\hat{x} = sign[\tanh(y)]$$

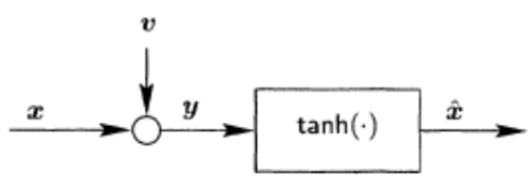


Figure 1.3. Optimal estimation of a BPSK signal embedded in unit-variance additive Gaussian noise.

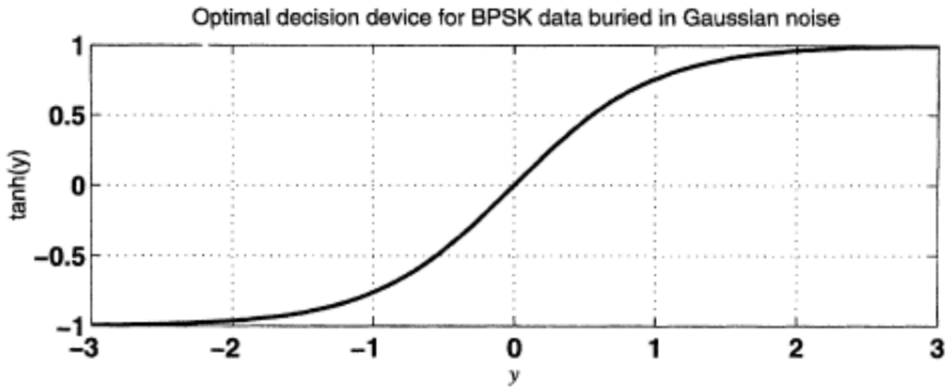


Figure 1.4. A plot of the function tanh(y).

As a note, from a designer perspective with a-priori knowledge that the symbols is a +1 or -1, this is equivalent to

$$\hat{x} = sign(y)$$

Note that the "optimal" estimator in this case can be quite complex, but the implementation of the concept significantly simplified.

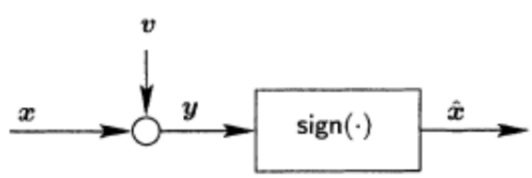


Figure 1.5. Sub-optimal estimation of a BPSK signal embedded in unit-variance additive Gaussian noise.

1.3.2 Orthogonality

If we make observation/estimate in a zero-mean, unit variance Gaussian environment:

$$y = x + v$$

The estimate of x given y from before is defined as

$$\hat{x} = E[x \mid y] = \int_{S_x} x \cdot f_{XY}(x \mid y) \cdot dx$$

Now we are interested in the difference

$$\tilde{x} \equiv x - \hat{x} = x - E[x \mid y]$$

For any function of y, g(y)

$$E_{x,y}[x \cdot g(y)] = \iint x \cdot g(y) \cdot f_{X,Y}(x,y) \cdot dx \cdot dy$$

$$E_{x,y}[x \cdot g(y)] = E_y[E_x[x \cdot g(y) \mid y]] = \int \left[\int x \cdot g(y) \cdot f_{XY}(x \mid Y = y) \cdot dx \right] \cdot f_Y(y) \cdot dy$$

$$E_{x,y}[x \cdot g(y)] = E_y[E_x[x \mid y] \cdot g(y)] = \int \left[\int x \cdot f_{XY}(x \mid Y = y) \cdot dx \right] \cdot g(y) \cdot f_Y(y) \cdot dy$$

Then,

$$E_{x,y}[\tilde{x} \cdot g(y)] = E_{x,y}[(x - \hat{x}) \cdot g(y)]$$

$$E_{x,y}[\tilde{x} \cdot g(y)] = \iint [x \cdot g(y) \cdot f_{X,Y}(x,y) - \hat{x} \cdot g(y) \cdot f_{X,Y}(x,y)] \cdot dx \cdot dy$$

$$E_{x,y}[\tilde{x} \cdot g(y)] = \iint [x \cdot g(y) \cdot f_{X,Y}(x,y) - \hat{x} \cdot g(y) \cdot f_{X,Y}(x,y)] \cdot dx \cdot dy$$

$$E_{x,y}[\tilde{x} \cdot g(y)] = E_y[E_x[x \mid y] \cdot g(y)] - \iint \hat{x} \cdot g(y) \cdot f_{X,Y}(x,y) \cdot dx \cdot dy$$

$$E_{x,y}[\tilde{x} \cdot g(y)] = E_y[E_x[x \mid y] \cdot g(y)] - \int \hat{x} \cdot g(y) \cdot \left[\int f_{X,Y}(x,y) \cdot dx \right] \cdot dy$$

$$E_{x,y}[\tilde{x} \cdot g(y)] = E_y[E_x[x \mid y] \cdot g(y)] - \int \hat{x} \cdot g(y) \cdot f_Y(y) \cdot dy$$

$$E_{x,y}[\tilde{x} \cdot g(y)] = E_y[E_x[x \mid y] \cdot g(y)] - \int E[x \mid y] \cdot g(y) \cdot f_Y(y) \cdot dy$$

$$E_{x,y}[\tilde{x} \cdot g(y)] = 0$$

Therefore, the error in the MSE estimate is “orthogonal” to all functions of the measurement y.

$$\tilde{x} \perp g(y)$$

As a further point, note that the estimate of x is a function of the measurement values of y. Therefore,

$$\tilde{x} \perp \hat{x}$$

From a probability sense, these relationships provide knowledge of the correlation of the values shown. As they are orthogonal, they must be uncorrelated!

"In summary, the optimal least-mean-square estimator is such that the estimation error is orthogonal to the estimator and, more generally, to any function of the observation." … "the Orthogonality condition is in fact a defining property of optimality in the least-mean-square sense."

Suboptimal estimators

It was previously shown that

$$\hat{x} = \tanh(y)$$

If we pick the suboptimal estimator

$$\hat{x} = sign(y)$$

We should be able to see that it is suboptimal if it violated the orthogonality condition. That is,

$$E[(x - sign(y)) \cdot sign(y)] \neq 0$$

This is homework problem 1.5 and the homework Matlab project 1.1.

1.3.3 Gaussian Random Variables

It is not always possible to determine a closed form expression for the optimal estimator

$$\hat{x} = E[x \mid y] = \int_{S_x} x \cdot f_{XY}(x \mid y) \cdot dx \ .$$

Therefore, we typically limit ourselves to linear or "affine" estimators based on

$$h(y) = \underline{A} \cdot y + \underline{B} \ .$$

See: http://en.wikipedia.org/wiki/Affinity_%28mathematics%29

In communications, we often have such a process where there is a scaling and offset in the value of a noisy signal. Notice that A and B above may be vectors!

Now: fun with vectors and matrices ….

The covariance matrix:

$$R = E\left[\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}\right) \cdot \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}\right)^T\right] \ .$$

$$R = E\begin{bmatrix} (x-\bar{x}) \cdot (x-\bar{x}) & (x-\bar{x}) \cdot (y-\bar{y}) \\ (y-\bar{y}) \cdot (x-\bar{x}) & (y-\bar{y}) \cdot (y-\bar{y}) \end{bmatrix} = R^T \ .$$

Every such matrix is: (1) symmetric $R=R^T$ and (2) non-negative definite (the determinate is 0 or positive det(R)>=0).

In terms of probability, the covariance matrix is

$$R = \begin{bmatrix} \sigma_x^{\ 2} & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^{\ 2} \end{bmatrix} = R^T \ .$$

The diagonals are the 2^{nd}-central moments and the off diagonals are the cross-correlations.

Jointly Gaussian

From the computation standpoint of a closed form solution … we prefer Gaussians due to the properties previously stated and the desire to "determine a closed form solution".

It makes it important to understand, (1) independent Gaussian noise and (2) Jointly Gaussian noise.

The joint Gaussian pdf is of the form

$$f_{X,Y}(x,y)=\frac{1}{\sqrt{2\pi}\cdot\det(R)}\cdot\exp\left(\frac{-1}{2}\cdot\begin{bmatrix}x-\bar{x} & y-\bar{y}\end{bmatrix}\cdot R^{-1}\cdot\begin{bmatrix}x-\bar{x}\\ y-\bar{y}\end{bmatrix}\right)$$

The individual variable pdf are as previously stated

$$f_X(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma_x}\cdot\exp\left(\frac{-\left(x-\bar{x}\right)^2}{2\cdot\sigma_x^{\;2}}\right),$$

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\cdot\sigma_y}\cdot\exp\left(\frac{-\left(y-\bar{y}\right)^2}{2\cdot\sigma_y^{\;2}}\right),$$

The conditional probability is then

$$f_{XY}(x\,|\,Y=y)=\frac{f_{XY}(x,y)}{f_Y(y)}$$

$$f_{XY}(x\,|\,Y=y)=\frac{\frac{1}{\sqrt{2\pi}\cdot\det(R)}\cdot\exp\left(\frac{-1}{2}\cdot\begin{bmatrix}x-\bar{x} & y-\bar{y}\end{bmatrix}\cdot R^{-1}\cdot\begin{bmatrix}x-\bar{x}\\ y-\bar{y}\end{bmatrix}\right)}{\frac{1}{\sqrt{2\pi}\cdot\sigma_y}\cdot\exp\left(\frac{-\left(y-\bar{y}\right)^2}{2\cdot\sigma_y^{\;2}}\right)}$$

Moving forward with linear algebra tricks: for a symmetric matrix R

$$R=\begin{bmatrix}A & B\\ B^T & C\end{bmatrix}=\begin{bmatrix}I & B\cdot C^{-1}\\ 0 & I\end{bmatrix}\cdot\begin{bmatrix}\left(A-B\cdot C^{-1}\cdot B^T\right) & 0\\ 0 & C\end{bmatrix}\cdot\begin{bmatrix}I & 0\\ C^{-1}\cdot B^T & I\end{bmatrix}.$$

See: http://en.wikipedia.org/wiki/Schur_complement

From before …

$$R = \begin{bmatrix} \sigma_x{}^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y{}^2 \end{bmatrix} = \begin{bmatrix} 1 & \sigma_{xy} / \sigma_y{}^2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Sigma & 0 \\ 0 & \sigma_y{}^2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ \sigma_{yx} / \sigma_y{}^2 & 1 \end{bmatrix}.$$

$$\Sigma^2 = \sigma_x{}^2 - \sigma_{xy}{}^2 / \sigma_y{}^2$$

Inverting by parts

$$R^{-1} = \begin{bmatrix} 1 & 0 \\ -\sigma_{xy} / \sigma_y{}^2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1/\Sigma^2 & 0 \\ 0 & 1/\sigma_y{}^2 \end{bmatrix} \cdot \begin{bmatrix} 1 & -\sigma_{yx} / \sigma_y{}^2 \\ 0 & 1 \end{bmatrix}.$$

Where

$$\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ -a & 1 \end{bmatrix}$$

Using

$$\det(R) = \det\begin{bmatrix} \sigma_x{}^2 & \sigma_{yx} \\ \sigma_{xy} & \sigma_y{}^2 \end{bmatrix} = \sigma_x{}^2 \cdot \sigma_y{}^2 - \sigma_{xy}{}^2 = \Sigma^2 \cdot \sigma_y{}^2$$

$$f_{XY}(x \mid Y = y) = \frac{1}{\sqrt{2\pi} \cdot \sqrt{\Sigma^2}} \cdot \exp\left(\frac{-\left[(x - \bar{x}) - \sigma_{xy} \cdot \sigma_y{}^{-2} \cdot (y - \bar{y})\right]^2}{2 \cdot \Sigma^2} \right)$$

In form, this can be related as an affine transformation of a single variable Gaussian pdf

$$f_Z(z) = \frac{1}{\sqrt{2\pi} \cdot \sigma_z} \cdot \exp\left(\frac{-(z - \bar{z})^2}{2 \cdot \sigma_z{}^2} \right),$$

Where the optimal estimator is

$$\hat{x} = E[x \mid y] = \int_{S_x} x \cdot f_{XY}(x \mid y) \cdot dx$$

or

$$\hat{x} = E[x \mid y] = \bar{x} + \sigma_{xy} \cdot \sigma_y{}^{-2} \cdot (y - \bar{y})$$

And the minimum mean-square error can be computed as

$$\tilde{x} \equiv x - \hat{x}$$

The minimum cost or variance becomes

$$E\left[\tilde{x}^2\right] = E\left[x^2 - \hat{x}^2\right] = \sigma_x^2 - \sigma_{\hat{x}}^2$$

$$E\left[\tilde{x}^2\right] = \Sigma^2 = \sigma_x{}^2 - {\sigma_{xy}}^2 \big/ {\sigma_y}^2$$

Note that the solution is strictly dependent on the means and variances of the original distribution.

See text for example 1.3.4

1.4 Estimation in the Complex and Vector Cases

Get used to complex sample values!

$$x = x_r + j \cdot x_i$$

Expected value

$$E[x] = E[x_r] + j \cdot E[x_i]$$

Variance

$$E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] = E\left[\left((x_r + j\cdot x_i) - (\bar{x}_r + j\cdot \bar{x}_i)\right)\cdot\left((x_r - j\cdot x_i) - (\bar{x}_r - j\cdot \bar{x}_i)\right)\right]$$

$$E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] = E\left[\begin{pmatrix} (x_r + j\cdot x_i)\cdot(x_r - j\cdot x_i) \\ -(x_r + j\cdot x_i)\cdot(\bar{x}_r - j\cdot \bar{x}_i) \\ -(\bar{x}_r + j\cdot \bar{x}_i)\cdot(x_r - j\cdot x_i) \\ +(\bar{x}_r + j\cdot \bar{x}_i)\cdot(\bar{x}_r - j\cdot \bar{x}_i) \end{pmatrix}\right] E\left[\begin{pmatrix} \left(x_r^{\ 2} + x_i^{\ 2}\right) \\ -2\cdot(x_r\cdot\bar{x}_r + x_i\cdot\bar{x}_i) \\ +\left(\bar{x}_r^{\ 2} + \bar{x}_i^{\ 2}\right) \end{pmatrix}\right]$$

$$E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] = E\left[x_r^{\ 2} + x_i^{\ 2}\right] - 2\cdot E\left[x_r\cdot\bar{x}_r + x_i\cdot\bar{x}_i\right] + \left(\bar{x}_r^{\ 2} + \bar{x}_i^{\ 2}\right)$$

$$E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] = E\left[x_r^{\ 2}\right] - \bar{x}_r^{\ 2} + E\left[x_i^{\ 2}\right] - \bar{x}_i^{\ 2}$$

$$\sigma_x^2 = \sigma_{x_r}^2 + \sigma_{x_i}^2$$

also

$$E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] = E\left[\begin{pmatrix} \left(x_r^{\ 2} - 2\cdot x_r\cdot\bar{x}_r + \bar{x}_r^{\ 2}\right) \\ +\left(x_i^{\ 2} - 2\cdot x_i\cdot\bar{x}_i + \bar{x}_i^{\ 2}\right) \end{pmatrix}\right] = E\left[\begin{pmatrix} (x_r - \bar{x}_r)^2 \\ +(x_i - \bar{x}_i)^2 \end{pmatrix}\right]$$

$$E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] == E\left[\left|(x_r + j\cdot x_i) - (\bar{x}_r + j\cdot\bar{x}_i)\right|^2\right]$$

$$E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] = E\left[\left|x-\bar{x}\right|^2\right]$$

Autocorrealtion

$$R_{xx} = E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right] = R_{xx}^{\ \ H} = E\left[(x-\bar{x})\cdot(x-\bar{x})^H\right]$$

This is called Hermitian symmetry

Cross correlation

In general, the values will be complex

$$E\left[(x-\bar{x})\cdot(y-\bar{y})^H\right] = E\left[\left(\left(x_r + j\cdot x_i\right) - \left(\bar{x}_r + j\cdot\bar{x}_i\right)\right)\cdot\left(\left(y_r - j\cdot y_i\right) - \left(\bar{y}_r - j\cdot\bar{y}_i\right)\right)\right]$$

$$E\left[(x-\bar{x})\cdot(y-\bar{y})^H\right] = \sigma_{xy}$$

Note that

$$E\left[(x-\bar{x})\cdot(y-\bar{y})^H\right] = \sigma_{xy} = \sigma_{xy}{}^H = E\left[(y-\bar{y})\cdot(x-\bar{x})^H\right]$$

Orthogonality

$$E\left[(x-\bar{x})\cdot(y-\bar{y})^H\right] = 0$$

Linear processes

$$y = A\cdot x$$

$$R_{yy} = E\left[(A\cdot x)\cdot(A\cdot x)^H\right] = E\left[(A\cdot x)\cdot\left(x^H\cdot A^H\right)\right]$$

$$R_{yy} = E\left[A\cdot x\cdot x^H\cdot A^H\right] = A\cdot E\left[x\cdot x^H\right]\cdot A^H$$

$$R_{yy} = A\cdot R_{xx}\cdot A^H$$

Note: what about for independent variables x and v

$$y = A\cdot x + B\cdot v$$

$$R_{yy} = A\cdot R_{xx}\cdot A^H + B\cdot R_{vv}\cdot B^H$$

1.4.3 Optimal Estimator in the Vector Case

Example 1.4.3: Various noisy measurements

$$y(i) = x + v(i)$$

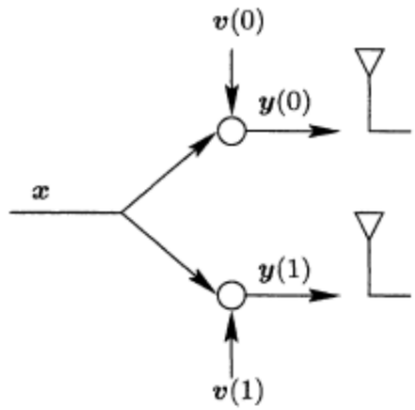


Figure 1.10. Reception by two antennas of a symbol x transmitted over two additive Gaussian-noise channels.

$$\begin{bmatrix} y(0) \\ y(1) \end{bmatrix} = x + \begin{bmatrix} v(0) \\ v(1) \end{bmatrix}$$

$$\hat{x} = h(\underline{y}) = h(y(0), y(1))$$

$$\tilde{x} = x - \hat{x}$$

The optimal estimator is still

$$\hat{x} = E(x \mid y(0), y(1))$$

For this case the extension from the previous solution is

$$\hat{x} = \tanh(y(0) + y(1))$$

In General

$$\begin{bmatrix} \hat{x}(0) \\ \\ \\ \hat{x}(p-1) \end{bmatrix} = \begin{bmatrix} h_0[y(0), y(1), \cdots, y(q-1)] \\ \\ \\ h_{p-1}[y(0), y(1), \cdots, y(q-1)] \end{bmatrix}$$

$$\underline{\tilde{x}} = \underline{x} - \underline{\hat{x}}$$

And we want

$$\min_{h_k(\ )}\{E[\tilde{x}(k)]\} \text{ for all k}$$

This is equivalent to

$$\min_{\{h_k(\ )\}}\{E[\underline{\tilde{x}}^H \cdot \underline{\tilde{x}}]\}$$

Tricks

$$E[\underline{\tilde{x}}^H \cdot \underline{\tilde{x}}] = \sum_{k=1}^{p-1} E[|\tilde{x}(k)|^2]$$

Or

$$E[\underline{\tilde{x}}^H \cdot \underline{\tilde{x}}] = Trace\{E[\underline{\tilde{x}} \cdot \underline{\tilde{x}^H}]\} = Trace\{R_{\tilde{x}\tilde{x}}\}$$

Example 1.4.4: Estimation of transmitted symbols

The input with independent, zero mean, unit norm Gaussian noise

$$\underline{y} = \underline{\underline{H}} \cdot \underline{x} + \underline{v}$$

$$\underline{\underline{H}} = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix}$$

$$y(0) = s(0) + v(0)$$

$$y(1) = 0.5 \cdot s(0) + s(1) + v(1)$$

Possible values for x, (for s = 1 with probability p, and s=-1 with prob. 1-p=q)

$$\underline{x} \in \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\} = \{ m_0, m_1, m_2, m_3 \}$$

Probabilities $\underline{x} \Rightarrow \{ p^2, q^2, p \cdot q, q \cdot p \}$

$$f_x(x) = \begin{matrix} p^2 \cdot \delta(x - m_0) \\ + q^2 \cdot \delta(x - m_1) \\ + p \cdot q \cdot \delta(x - m_2) \\ + q \cdot p \cdot \delta(x - m_3) \end{matrix}$$

For the noise (note the covariance matrix is an identity matrix)

$$f_v(v) = \frac{1}{2\pi} \cdot \exp\left(\frac{-v^t \cdot v}{2} \right)$$

Then

$$f_{x,y}(x, y) = f_x(x) \cdot f_v(y - H \cdot x)$$

$$f_{x,y}(x, y) = \begin{matrix} p^2 \cdot f_v(y - H \cdot m_0) \cdot \delta(x - m_0) \\ + q^2 \cdot f_v(y - H \cdot m_1) \cdot \delta(x - m_1) \\ + p \cdot q \cdot f_v(y - H \cdot m_2) \cdot \delta(x - m_2) \\ + q \cdot p \cdot f_v(y - H \cdot m_3) \cdot \delta(x - m_3) \end{matrix}$$

Determining pdf of y based on x and v

$$f_y(y) = \begin{array}{l} p^2 \cdot f_v(y - H \cdot m_0) \\ + q^2 \cdot f_v(y - H \cdot m_1) \\ + p \cdot q \cdot f_v(y - H \cdot m_2) \\ + q \cdot p \cdot f_v(y - H \cdot m_3) \end{array}$$

$$\hat{x} = E[x \mid y] = \int_{S_x} x \cdot f_{XY}(x \mid y) \cdot dx \ .$$

$$f_{XY}(x \mid y) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

Calculating Hm

$$H \cdot m_0 = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix} \quad H \cdot m_1 = \begin{bmatrix} -1 \\ -1.5 \end{bmatrix} \quad H \cdot m_2 = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix} \quad H \cdot m_2 = \begin{bmatrix} -1 \\ 0.5 \end{bmatrix}$$

Calculating $v^T v$

$$(y - H \cdot m_0)^T (y - H \cdot m_0) \Rightarrow (y(0) - 1)^2 + (y(1) - 1.5)^2$$

$$(y - H \cdot m_1)^T (y - H \cdot m_1) \Rightarrow (y(0) + 1)^2 + (y(1) + 1.5)^2$$

$$(y - H \cdot m_2)^T (y - H \cdot m_2) \Rightarrow (y(0) - 1)^2 + (y(1) + 0.5)^2$$

$$(y - H \cdot m_3)^T (y - H \cdot m_3) \Rightarrow (y(0) + 1)^2 + (y(1) - 0.5)^2$$

Let

$$a = p^2 \cdot \frac{1}{2\pi} \cdot \exp\left(- \frac{(y(0) - 1)^2 + (y(1) - 1.5)^2}{2} \right)$$

$$b = q^2 \cdot \frac{1}{2\pi} \cdot \exp\left(- \frac{(y(0) + 1)^2 + (y(1) + 1.5)^2}{2} \right)$$

$$c = p \cdot q \cdot \frac{1}{2\pi} \cdot \exp\left(- \frac{(y(0) - 1)^2 + (y(1) + 0.5)^2}{2} \right)$$

$$d = p \cdot q \cdot \frac{1}{2\pi} \cdot \exp\left(- \frac{(y(0) + 1)^2 + (y(1) - 0.5)^2}{2} \right)$$

Then

$$\hat{x} = \frac{m_0 \cdot a + m_1 \cdot b + m_2 \cdot c + m_3 \cdot d}{a + b + c + d} \quad .$$

Further simplification is possible, but I've run out of time ….

1.4.5: Spherically Invariant Gaussian Variables

The results are important, so please read this section. You will see the resulting equation again in another context.

1.A Hermitian and Positive-Definite Matrices

Hermitian Matrix Properties

Spectral decomposition

Taking the eigenvalues

$$\lambda_i \cdot u_i = A \cdot u_i$$

Premultiply

$$u_i^{\ H} \cdot \lambda_i \cdot u_i = \lambda_i \cdot \|u_i\|^2 = u_i^{\ H} \cdot A \cdot u_i$$

And

$$\left(\lambda_i \cdot \|u_i\|^2\right)^H = \left(u_i^{\ H} \cdot A \cdot u_i\right)^H$$

$$\left(u_i^{\ H} \cdot A \cdot u_i\right)^H = u_i^{\ H} \cdot A^H \cdot u_i = u_i^{\ H} \cdot A \cdot u_i$$

Therefore,

$$conj(\lambda_i) \cdot \|u_i\|^2 = u_i^{\ H} \cdot A \cdot u_i = \lambda_i \cdot \|u_i\|^2$$

So the eigenvalues are purely real

$$\lambda_i = conj(\lambda_i)$$

The eigenvalues and eigenvectors can be represented in the form of a matrix

$$A = U \cdot \Lambda \cdot U^H$$

where Λ is a diagonal matrix of the eigenvalues and U is a matrix who’s columns are the eigenvector (in order of the diagonals eigenvalues).

Positive definite

$$\lambda_i \cdot u_i = A \cdot u_i \text{, with } \|u_i\|^2 = 1$$

Premultiply

$$u_i^H \cdot \lambda_i \cdot u_i = \lambda_i \cdot \|u_i\|^2 = u_i^H \cdot A \cdot u_i = \lambda_i$$

Taking

$$x^H \cdot A \cdot x = x^H \cdot U \cdot \Lambda \cdot U^H \cdot x$$

Forming a square-root matrix of the Λ term

$$x^H \cdot A \cdot x = x^H \cdot U \cdot \Lambda^{1/2} \cdot \Lambda^{1/2} \cdot U^H \cdot x$$

Let

$$y = \Lambda^{1/2} \cdot U^H \cdot x$$

Then

$$x^H \cdot A \cdot x = x^H \cdot U \cdot \Lambda^{1/2} \cdot \Lambda^{1/2} \cdot U^H \cdot x = y^H \cdot y = \|y\|^2$$

Therefore,

$$x^H \cdot A \cdot x = \|y\|^2 > 0$$

Rayleigh-Ritz characterization of eigenvalues

$$\lambda_{\min} \cdot \|x\|^2 \le x^H \cdot A \cdot x \le \lambda_{\max} \cdot \|x\|^2$$

Rayleigh-Ritz ratio

$$\lambda_{\min} \le \left(\frac{x^H \cdot A \cdot x}{x^H \cdot x} \right) \le \lambda_{\max}$$

Thererfore

$$\lambda_{\min} = \min_{\|x\|^2=1} \left(x^H \cdot A \cdot x \right)$$

$$\lambda_{\max} = \max_{\|x\|^2=1} \left(x^H \cdot A \cdot x \right)$$

Necessary and sufficient conditions for A to be positive definite.

(I) $x^T \cdot A \cdot x > 0$ for all nonzero vectors x.

(II) All eigenvalues of A satisfy $\lambda_i > 0$.

(III) All submatricies of A have positive determinants.

(IV) All the pivots (without row exchange) satisfy $d_i > 0$..

(V) There exists a nonsingular matrix W such that $A = W^T \cdot W$.
(Note: this is called a Cholesky decomposition and W is upper triangular).

from G. Strang, “Linear Algebra and Its Applications”, 2nd ed., Academic Press, 1980.

Summary of Main Results

When faced with the estimation problem where multiple measurements in the presence of noise or other statistically define process

$$y = x + v$$

and we which to determine the minimum mean square estimate of x or $\hat{x}$. Where the error in the estimate is defined as

$$\tilde{x} \equiv x - \hat{x}$$

Define a cost function, J, based on the error mean and variance

$$J[\tilde{x}] = A \cdot E[\tilde{x}^2] + B \cdot E[\tilde{x}]$$

If ***no observations are available***:

$$\hat{x} = \bar{x}$$

When the mmse is selected,

$$J[\tilde{x}] = A \cdot \sigma_x^2 + B \cdot E[x - \hat{x}] = A \cdot \sigma_x^2$$

If ***dependent observations are available***:

$$\hat{x} = h(y) = E[x \mid y] = \int_{S_x} x \cdot f_{XY}(x \mid y) \cdot dx$$

where the integral denotes the domain of x.

The minimum cost or variance becomes

$$E[\tilde{x}^2] = E[x^2 - \hat{x}^2] = \sigma_x^2 - \sigma_{\hat{x}}^2$$

Orthogonality

$$E_{x,y}[\tilde{x} \cdot g(y)] = E_{x,y}[(x - \hat{x}) \cdot g(y)] = 0$$

$$\tilde{x} \perp g(y)$$

$$\tilde{x} \perp \hat{x}$$

Circular Gaussian Example

The estimate is completely determined from knowledge of the first and second moments … the means, covariances, and cross-covariances

Real vs Complex

While many examples are describe using real numbers or real numbered processes, the complex derivations can be performed.